DOCUMENT RESUME

ED 441 009 TM 030 807

AUTHOR	Rogers, Bruce G.
TITLE	Toward "Constructing" the Concept of Statistical Power: An Optical Analogy.
PUB DATE	2000-04-00
NOTE	6p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	College Students; *Concept Formation; Higher Education; *Statistics; *Teaching Methods
IDENTIFIERS	*Analogy; *Power (Statistics)

ABSTRACT

This paper presents a visual analogy that may be used by instructors to teach the concept of statistical power in statistical courses. Statistical power is mathematically defined as the probability of rejecting a null hypothesis when that null is false, or, equivalently, the probability of detecting a relationship when it exists. The analogy involved a group of hikers in desert heat who are faced with the possibility that a pool of water seen by only one hiker is a mirage. Effect size, sample size, and significance level are discussed in terms of the mirage analogy. Type I errors, error variance, and directionality are also discussed in terms of the optical analogy. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 441 009

Toward "Constructing" the Concept of Statistical Power: An Optical Analogy

Bruce G. Rogers

University of Northern Iowa

Roundtable Paper presented at the
April 2000 Annual Meeting of the
American Educational Research Association
New Orleans

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

B. Rogers

1

Address correspondence to Dr. Bruce Rogers, Department of Educational Psychology, University of Northern Iowa, Cedar Falls, IA 50614-0607 Bruce.Rogers@UNI.edu

TM030807

BEST COPY AVAILABLE

The concept of the statistical power of a test is an abstract notion which is found to be very challenging for many students in both the introductory and advanced courses in educational statistics. One of the reasons frequently cited by students is the difficulty of visual conception. In former years, power graphs were printed in some texts, but gaining a visual conception of the abstraction was often challenging to the student. It is the purpose of this paper to present a visual analogy that may be employed by instructors to teach this concept to their students in statistical courses. It is anticipated that the analogy will then be useful to students in helping them to construct, in their own minds, the concept of statistical power.

Statistical power is mathematically defined as the probability of rejecting a null hypothesis when that null is false, or, equivalently, as the probability of detecting a relationship when it exists. In 1969, Jacob Cohen published his well-known book devoted to the topic of statistical power analysis. Since the appearance of that book (which was subsequently revised in 1988), the terms "statistical power" and "effect size" have been introduced into almost all introductory statistics textbooks, the tables have been widely used, and more recently, some computer programs have been written for ease of estimating power and desired sample sizes. Since students will be expected to become familiar with this topic, instructors may facilitate that process with an approach toward helping students construct this concept is their own minds. As Elliott Eisner (1999) explains it, what students "make of what we provide is a function of what they construct from what we offer. Meanings are not given, they are made." (p. 658)

The term "constructivism" has been given a variety of meanings in the educational literature. The concept of "scientific constructivism"(see an explication by Michael Battista (1999)) conveys the meaning used in the context of this paper. Scientists construct models, such as the familiar Bohr atom, as one means of understanding atomic structure. In that model, electrons are envisioned as planets. It is important to stress that a model is what we create, i.e., 'construct', in our own minds to help us understand observable phenomena. That model is always incomplete, therefore, we may construct other models to help us visualize the phenomena in another manner. Thus, the atomic structure may also be conceived with enveloping clouds representing electrons. Both models can be of value to students as they seek understanding.

Likewise, the instructor in a statistics course must offer materials for helping students to construct, in their own minds, meaningful ways to interpret the concept of statistical power and how it relates to other important statistical concepts. Craig Enders (1999) provided an excellent analogy in which he compared power to the likelihood of detecting a radio signal. Since a multitude of such examples may help the student in their construction of the meaning of statistical power, a visual analogy is offered in this paper.

An Explication of the Analogy

The probability concept of statistical power will now be developed using an analogy involving a group hiking in a desert on a hot summer day. It is noontime, time to stop to eat lunch. But one of the hikers reports seeing a pool of water in the distance and suggests that they

eat their lunch by the cool water. Their map and guide book suggests that at certain times of the year pools of water have existed following a rain storm, but there is no way for the hikers to know for sure whether a water pool exists or not at the present time. In their previous experiences, the hikers have found that whenever someone claims to see a pool of water, there is a possibility that the person has been deceived by a mirage. In this context, power is the probability, or likelihood, that the hikers will make the right decision when a pool of water actually exists.

Parameters that Influence Statistical Power

It is clear that several things can influence their ability to distinguish between actual water and a mirage. Textbooks usually describe three factors which influence statistical power, namely, effect size, sample size and significance level. In addition, error variance and directional tests are usually discussed. In the remainder of this paper, these concepts will be developed in the context of the optical analogy.

Effect Size. In this discussion, effect size might be considered analogous to the size of the pool of water. .A large pool of water is more likely to be distinguished from a mirage than is a medium sized pool or a small pool. If the pool is small enough, it might be considered to be inconsequential, essentially the same as if no water existed. Can the hikers agree on what is meant by the terms 'small pool', 'medium pool' and 'large pool'? The late Jacob Cohen (1988) proposed a set of conventions for 'small', 'medium' and 'large' effect sizes, which has come to be accepted as reasonable by other educational psychologists. It is clear that the effect size, represented by the pool size in this analogy, is one of the factors that affects the likelihood of detecting the phenomena. The size of the pool cannot be manipulated by the hiker. Therefore, in addition to declaring that a pool was sighted, it is important for the hiker to estimate the size of the pool. Bruce Thompson (1999) has described the strong effort that is being made to encourage all educational statisticians to report the effect size for any observed phenomena in research.

Sample Size. In this analogy, sample size might be considered as the number of times that the image of the pool is seen. It could be reported as seen by four of the hikers. We would call this an 'independent sample', assuming that the hikers did not consult with each other prior to making their own sighting. Or, one hiker might have sighted it four times from different positions. We would call this a 'dependent sample' or 'repeated sample'. As the number of observations increases, the confidence of the hikers can increase. For a small pool, more observations will be necessary to obtain a given level of confidence than would be necessary for a medium pool or a large pool. Clearly, the sample size is another one of the factors that affects the likelihood, or power, of detecting the phenomena.

Significance Level. How much risk is one of the hikers willing to take before telling

someone else that a pool of water has been detected? The hiker must decide this level of risk prior to making a conclusion. Above this level of risk, the image will be dismissed as a mirage, while a visual impression below this level of risk will be announced to others. If the hiker sets the level of risk too low, then it is possible that an existing pool will not be reported, that is, the power of the test will be reduced. If the hiker sets the level of risk too high, then the reporting of mirages may lead to wasted effort on the part of others. Perhaps one hiker will be willing to risk making an error one out of twenty times, while another hiker would feel comfortable with making an error only one of a hundred times. Whatever the level of risk the hiker chooses, that level should be stated prior to the reporting of the sighting of a pool. If the number of sights can be increased, then the likelihood (power) of detecting a pool of water can be increased with the same level of risk.

In statistical parlance, this level of risk is called Type I error, because it is usually the first type of error to be considered. Those who compile statistical tables have commonly chosen values of .05, .01, and .001. With the advent of the computer, it is computationally feasible to estimate the exact likelihood that a given sample could have occurred completely by chance. This is commonly reported as a p value. However, the researcher must still decide the level of risk that will be used and that will directly affect the power of the test. As Type I error is reduced, power is reduced. Similarly, as power is increased, type I error is increased. If it is possible to increase the sample size, then the Type I error can be held constant and the power will be increased.

Error variance. One source of errors that may confound the decision of a hiker arises from the influences that tend to obscure the vision of the hiker. Smudged eyeglasses, dust in the atmosphere and astigmatism in the eye are examples of the many influences that may affect clarity of vision. The angle of the rays of the sun and the surface features of the ground will also affect the appearance of a mirage. These potentially have the effect of both reducing the power of the test and increasing the probability of a Type I error.

All statistics are based upon the process of measurement, which is the assigning of numbers to observed phenomena. The educational statistician must always be aware of this source of variation. In educational studies, it often includes the reliability of a test, the validity of a procedure, and the inter-rater agreement of observations. All of these clearly have the effect of reducing the power of detecting an effect that exists.

Directionality. As the hikers search for a pool of water, should they concentrate only to the left of the trail, only to the right of the trail, or should they look on both sides of the trail? In weighing the options, the hikers see advantages and disadvantages. If they look only to the left, they can focus their attention on a smaller area and thus reduce the risk of mistaking a mirage for an actual pool. However, if a pool exists on the right, they will never see it. On the other hand, if they look both to the left and to the right, they can only look in one direction one-half of the time, and thus the risk of making an error will increase. If the possibility of a mirage appears to be a random event, then it would appear that it could occur equally likely on either side. If it is known that a mirage cannot occur on the right-hand side, then the nature of the problem is changed and

should be reformulated.

Those who favor one-tailed tests sometime justify it by saying that the research hypothesis is directional. But if it is already known that the phenomena can only be in one direction, then there is no need for a statistical test. The statistical test is based upon the assumption of random variation, which leads to the use of probability theory. Under that assumption, it is equally likely that observations can fall in either tail. Thus, any user of a directional test will need to be prepared to defend the logical and mathematical questions that will arise. In a non-directional test, the power of the test can be estimated directly, while in a directional test, the power of the test is problematical.

Educational Importance

For many students in a statistics course, the concept of statistical power is a new and abstract concept. Textbook authors attempt to clarify the concept with various drawings which may be mathematically correct but which may be difficult for the beginning student to comprehend. A visual analogy like the above may help students begin to construct, in their own minds, an interpretation of power that is meaningful to them and which they can then use to apply as they read mathematical explanations to further construct their understanding of the concept of statistical power. As they develop in their understanding of this topic, the mathematical drawings of overlapping curves and the sweeping power curves will begin to be less confusing and more enlightening to them.

References

Battista, M. T. (1999, Feb.). The mathematical miseducation of America's youth. Phi Delta Kappan, 80(6) 425-433.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). NY: Academic Press.

Eisner, E. W. (1999, May). The uses and limits of performance testing. Phi Delta Kappan, 80(9) 658-660.

Enders, C. K. (1999, April). Teaching statistical power: Providing a conceptual overview using an analogy. Paper presented at the 1999 Annual Meeting of the AERA.

Thompson, B. (1999, April). Common methodology mistakes in educational research. Paper presented at the 1999 Annual Meeting of the AERA. ERIC # ED429110.